



## RESEARCH

## Open Access

# CRISPR transcript processing: a mechanism for generating a large number of small interfering RNAs

Marko Djordjevic<sup>1\*</sup>, Magdalena Djordjevic<sup>2</sup> and Konstantin Severinov<sup>3,4,5</sup>

## Abstract

**Background:** CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated sequences) is a recently discovered prokaryotic defense system against foreign DNA, including viruses and plasmids. CRISPR cassette is transcribed as a continuous transcript (pre-crRNA), which is processed by Cas proteins into small RNA molecules (crRNAs) that are responsible for defense against invading viruses. Experiments in *E. coli* report that overexpression of *cas* genes generates a large number of crRNAs, from only few pre-crRNAs.

**Results:** We here develop a minimal model of CRISPR processing, which we parameterize based on available experimental data. From the model, we show that the system can generate a large amount of crRNAs, based on only a small decrease in the amount of pre-crRNAs. The relationship between the decrease of pre-crRNAs and the increase of crRNAs corresponds to strong linear amplification. Interestingly, this strong amplification crucially depends on fast non-specific degradation of pre-crRNA by an unidentified nuclease. We show that overexpression of *cas* genes above a certain level does not result in further increase of crRNA, but that this saturation can be relieved if the rate of CRISPR transcription is increased. We furthermore show that a small increase of CRISPR transcription rate can substantially decrease the extent of *cas* gene activation necessary to achieve a desired amount of crRNA.

**Conclusions:** The simple mathematical model developed here is able to explain existing experimental observations on CRISPR transcript processing in *Escherichia coli*. The model shows that a competition between specific pre-crRNA processing and non-specific degradation determines the steady-state levels of crRNA and is responsible for strong linear amplification of crRNAs when *cas* genes are overexpressed. The model further shows how disappearance of only a few pre-crRNA molecules normally present in the cell can lead to a large (two orders of magnitude) increase of crRNAs upon *cas* overexpression. A crucial ingredient of this large increase is fast non-specific degradation by an unspecified nuclease, which suggests that a yet unidentified nuclease(s) is a major control element of CRISPR response. Transcriptional regulation may be another important control mechanism, as it can either increase the amount of generated pre-crRNA, or alter the level of *cas* gene activity.

**Reviewers:** This article was reviewed by Mikhail Gelfand, Eugene Koonin and L Aravind.

**Keywords:** CRISPR/Cas, Transcript processing, Small RNA, CRISPR expression regulation, CRISPR/Cas response

\* Correspondence: [dmarko@bio.bg.ac.rs](mailto:dmarko@bio.bg.ac.rs)

<sup>1</sup>Institute of Physiology and Biochemistry, Faculty of Biology, University of Belgrade, Belgrade, Serbia

Full list of author information is available at the end of the article

## Background

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) cassettes are present in almost every known archaeal genome and in about half of the known bacterial genomes [1-3]. A CRISPR cassette consists of identical direct repeats of about 30 bp in length, interspaced with spacers of similar length [4]. The length of different spacers within the same cassette is the same, while sequences of these spacers are different. In many organisms, these spacer sequences closely match sequences of bacteriophages (bacterial viruses) infecting this or closely related organisms [5-7]. It was recently discovered that CRISPR/Cas loci function as an adaptive immunity system, which is responsible for defending prokaryotic cell against viruses and plasmids [8,9]. A match between a CRISPR spacer and sequence in invading DNA provides immunity to infection [5-9].

In *E. coli*, promoters that transcribe CRISPR cassettes and *cas* genes are distinct, and are (at least under normal growth conditions) considered to be poorly active due to repression by H-NS transcription factor [10]. The entire CRISPR cassette is transcribed as a long continuous transcript [10,11], which is then processed by one of the Cas proteins (CasE), to small RNA molecules (crRNAs) [11,12]. Once crRNAs are generated, they bind a large multisubunit complex of Cas proteins called Cascade and target it to matching DNA of viruses and plasmids, ultimately leading to its destruction [13].

While it is clear that CRISPR/Cas system in *E. coli* is functional [11,14], virus infection in itself appears not to lead to system induction (at least under normal conditions) [15], and physiological conditions under which the system is induced yet have to be determined [13]. Consequently, functioning of this system has been investigated by either artificial overexpression of *cas* genes and CRISPR array from plasmids, or by inhibition of H-NS repression of *cas* and CRISPR promoters [11,12,16]. In a recent study, *cas* genes were overexpressed in *E. coli*, and resulting changes in the levels of pre-crRNAs and crRNAs were quantitatively measured [11]. In cells with endogenous (uninduced) *cas* expression, the abundance of pre-crRNA and individual crRNAs was low, below 10 molecules per cell. When CasE was overexpressed, the abundance of crRNAs increased dramatically, to about 1000 molecules per cell, while pre-crRNA became undetectable. There is, therefore, a large (at least two orders of magnitude) increase in abundance of individual crRNAs, accompanied by a much smaller (less than tenfold) decrease of pre-crRNA. It remains unclear if (and by what model) this strong amplification of crRNA upon *cas* overexpression can be explained. Answering this question is a major goal of this paper.

Furthermore, the experiments discussed above correspond to measurements where *cas* genes and CRISPR

arrays are overexpressed to a fixed level [10-12,16]. On the other hand, it is important to explore how changes of the relevant parameters affect generation of crRNAs, since such understanding can provide important clues about the mechanism of the endogenous system induction. Finally, the available experiments correspond to steady-state measurements of transcript amounts, i.e. come from measurements taken long after *cas* genes overexpression has been induced. However, the steady-state regime may not be directly relevant for system function under natural conditions, where the amount of generated crRNA immediately after system induction (i.e., for example, after virus infection) may be more relevant. While it is hard to experimentally assess either different levels of parameter changes or kinetics of the transcript accumulation, this analysis can be readily done through mathematical modeling, which is another major goal of this paper.

We will in this paper present a simple mathematical model of CRISPR expression that is able to *i)* determine biochemical parameters relevant for CRISPR transcript processing, *ii)* explain the observed large amplification of crRNAs, *iii)* assess how different levels of change in the transcription and processing rates affect steady-state levels and kinetics of crRNA accumulation.

## Results

### Model definition

In this section, we will propose a simple model of CRISPR transcript processing. The model is in accordance with the following experimental observations:

- i) Endogenous (uninduced) levels of pre-crRNAs and crRNAs are low (~10 copies per cell) [11,12,16], which was reported to be a consequence of repression of *cas* and (to a smaller extent) CRISPR promoters by H-NS [10].
- ii) One of the Cas proteins (CasE) is responsible for processing pre-crRNAs to crRNAs [11,12]. When CasE is overexpressed, the amount of crRNAs increases for about two orders of magnitude, while the amount of pre-crRNAs drops to only few transcripts per cell [11]. Overexpression of CasE affects only the processing rate of pre-crRNA to crRNA, since it has been shown [11] that CasE does not influence either pre-crRNA transcription rate or crRNA stability.
- iii) In addition to being processed by CasE, pre-crRNA is also degraded by an unspecified nuclease [10,11]. As a consequence of this degradation, pre-crRNA decays with a half-life of ~1 min without generating crRNAs. On the other hand, crRNAs are observed to be much more stable [11].
- iv) It is currently unclear how CRISPR/Cas system is induced under natural conditions [13]. It was,

however, showed that the repression of the *cas* promoter by H-NS can be relieved by a transcription activator (LeuO) [16]. It was consequently proposed that the endogenous system induction may involve activation of *cas* and (to a smaller extent) CRISPR promoters, through abolishment of H-NS repression [10].

The simplest model of CRISPR transcript processing, which is in accordance with the experimental observations summarized above, is schematically shown in Figure 1. In the scheme, we denote concentrations of the unprocessed (pre-crRNA) and processed (crRNA) transcripts as, respectively,  $[u]$  and  $[p]$ . The unprocessed transcripts (pre-crRNAs) are transcribed with rate  $\phi$ ; pre-crRNAs are further either non-specifically degraded with rate  $\lambda_u$ , or processed by CasE with rate  $k$ . By non-specific degradation, we mean degradation that does not lead to accumulation of crRNA. Processing of pre-crRNA by CasE leads to formation of individual crRNAs, which are further degraded with rate  $\lambda_p$ . Based on the experimental results [11], we take  $\lambda_u \sim 1 \text{ min}^{-1}$ ,  $\lambda_p \sim 1/100 \text{ min}^{-1}$ , and  $[u] \sim [p] \sim 10$ .

While the uninduced values of pre-crRNA transcription and processing rates ( $\phi$  and  $k$ ) have not been experimentally measured, they can be determined from equations that describe kinetics of the system in Figure 1 (see the next section). When the system is induced, both  $k$  and  $\phi$  can be increased. Since CasE is solely responsible for processing of pre-crRNA to crRNA, the value of the processing rate  $k$  depends on the amount of CasE. Consequently, the increase of  $k$  is due to increased amount of CasE, which is a consequence of a larger transcription activity of *cas* promoters. Similarly,  $\phi$  can be increased if the CRISPR promoter becomes more active.

In the next subsection, we will show that the simple model, schematically shown in Figure 1, together with

experimentally inferred parameter values summarized above, can indeed explain the observed large crRNA amplification upon induction of *cas* gene expression. We will afterwards explore kinetics of crRNA generation, and investigate how modulation of pre-crRNA transcription and processing rate ( $\phi$  and  $k$ ) affects generated crRNA amounts.

#### Uninduced system parameters

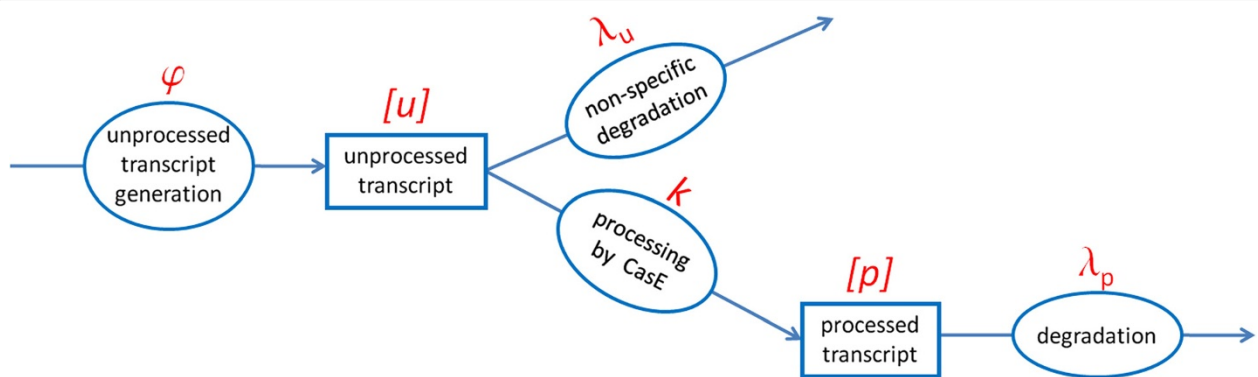
Starting from equations that describe the system kinetics (see Methods), it is straightforward to obtain expressions for uninduced values of pre-crRNA transcription and processing rates ( $\phi$  and  $k$ ):

$$\phi = \lambda_u [u] + \lambda_p [p], \quad (0.1)$$

$$k = \lambda_p \frac{[p]}{[u]} \quad (0.2)$$

In the equations above  $[u]$  and  $[p]$  are, respectively, (uninduced) steady state amounts of pre-crRNA and crRNA, while  $\lambda_u$  and  $\lambda_p$  are defined in Figure 1.

By using the numerical values stated in the previous section, from Eq. (0.1) we obtain  $\phi \sim 10\lambda_u \sim 10 \text{ min}^{-1}$ . This value corresponds to a moderately strong transcription activity; note that transcription activity of very strong rRNA promoters is  $\sim 60 \text{ min}^{-1}$ , while basal activity of a very weak uninduced  $\lambda_{\text{PRM}}$  promoter is  $\sim 1/7 \text{ min}^{-1}$  [17]. It is interesting that in experimental studies the CRISPR promoter was labeled as weak, based on measured small amount of pre-crRNA [10,11]. The small amount of pre-crRNA is actually a consequence of a high non-specific decay rate of pre-crRNA (note that pre-crRNA half life is  $\sim 1 \text{ min}$ ), which has to be matched by the relatively high activity of the CRISPR promoter. The moderately high transcription rate of the CRISPR promoter implies a weak repression of this promoter by H-NS, which is consistent with the experimental finding that repression



**Figure 1 Model of CRISPR transcript processing.** The unprocessed transcripts (pre-crRNAs) are generated with rate  $\phi$ , and are consequently either (non-specifically) degraded with rate  $\lambda_u$ , or processed to crRNAs by CasE with rate  $k$ . crRNAs are then degraded with rate  $\lambda_p$ . Concentrations of pre-crRNAs (unprocessed transcripts) and crRNAs (processed transcripts) are denoted as, respectively,  $[u]$  and  $[p]$ .

of the CRISPR promoter by H-NS is significantly weaker compared to the repression of the *cas* promoter [10,16].

Similarly, by using numerical values from the previous subsection and Eq. (0.2), we obtain  $k \sim \lambda_p \sim 1/100 \text{ min}^{-1}$ . Therefore, pre-crRNA to crRNA processing rate ( $k$ ) is an order of magnitude smaller than pre-crRNA decay rate ( $\lambda_u$ ). Due to this, when the system is uninduced, almost all generated pre-crRNA is rapidly degraded (see Figure 1), which results in small crRNA amounts, despite the moderately high transcription rate ( $\phi$ ) of the uninduced promoter. As we will show in the next subsection, when the system is induced and  $k$  is increased, the system switches from the state in which almost all of the generated pre-crRNA is degraded, to the state in which most of the generated pre-crRNA is processed to crRNA.

### Overexpression of *cas* genes

We next analyze the experiments in which CasE is overexpressed, and the transcript numbers are quantified [11]. In these experiments, the number of pre-crRNA and crRNA transcripts has been measured both before and after the system induction. In the analysis below, we assume that overexpression of CasE leads to an increase of pre-crRNA to crRNA processing rate from  $k$  to  $k'$ , while it has been experimentally shown that the rest of the parameters remain unchanged (see above). Furthermore, we denote pre-crRNA and crRNA amounts upon CasE overexpression as, respectively,  $[u]'$  and  $[p]'$ . Note that primes in our notation correspond to the quantity values after the system induction, rather than to derivatives.

We aim to understand the large amplification of crRNA, where, upon CasE overexpression, a decrease from about  $\sim 10$  pre-crRNA transcripts present in uninduced cells leads to about two orders of magnitude increase in the amount of crRNA ( $\sim 1000$  transcripts). To that end, it is useful to derive a relationship between the changes in the number of crRNAs ( $\Delta[p] \equiv [p]' - [p]$ ) and pre-crRNAs ( $\Delta[u] \equiv [u]' - [u]$ ). By using the equations for the system kinetics (see Methods), one can derive the following (exact) relation:

$$\Delta[p] = -\frac{\lambda_u}{\lambda_p} \Delta[u] \quad (0.3)$$

Note that the minus sign indicates that the decrease in the number of unprocessed transcripts (pre-crRNA), leads to an increase in the number of processed transcripts (crRNA). From the relationship above follows that the crRNA increase is directly proportional to the pre-crRNA decrease, where the constant of proportionality is equal to 100 ( $\lambda_u/\lambda_p \sim 100$  - see the previous section). This large constant of proportionality in Eq. (0.3)

explains the experimentally observed large amplification of crRNA upon CasE overexpression. That is, according to Eq. (0.3),  $\sim 10$  molecule decrease in pre-crRNA ( $\Delta[u] \sim 10$ ), leads to two orders of magnitude larger increase in crRNA ( $\Delta[p] \sim 1000$ ), as observed in the experiments. Therefore, Eq. (0.3) shows that the system acts as a strong linear amplifier, where the increase of crRNA is directly proportional to the decrease of pre-crRNA, and where a small number of pre-crRNAs are amplified to a large number of crRNAs.

Experiments also report that, upon Cas overexpression, the amount of pre-crRNA decreases for about one order of magnitude, which allows estimating the extent of increase of pre-crRNA to crRNA processing rate ( $k$ ). From equations that describe the system kinetics (see Methods), it is straightforward to show that the relative decrease of pre-crRNA amount is given by

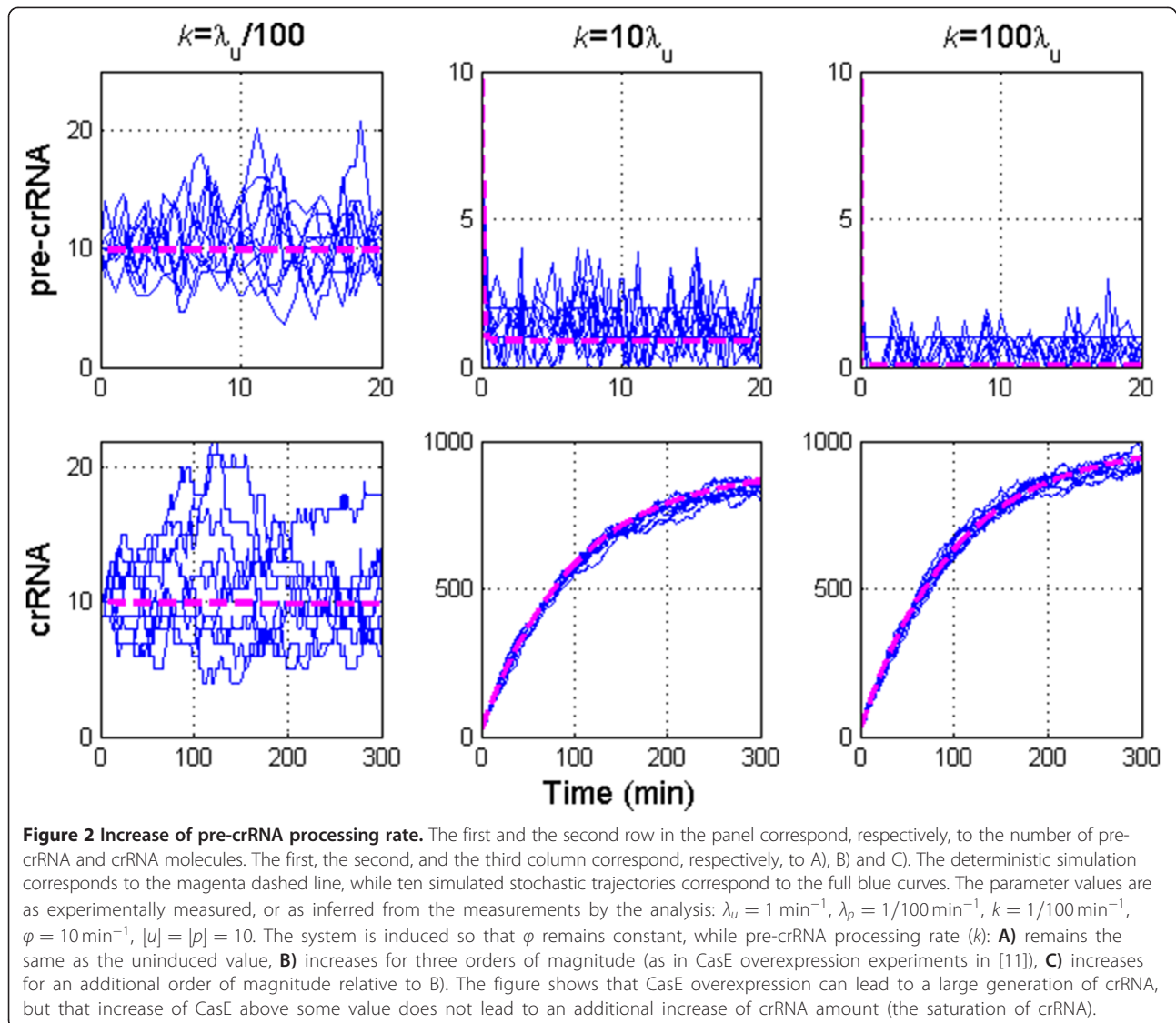
$$\frac{[u]}{[u]'} = \frac{\lambda_u + k'}{\lambda_u + k}. \quad (0.4)$$

It is experimentally observed that  $[u]/[u]' \sim 10$ , so from Eq. (0.4) follows  $k' \sim 10(\lambda_u + k)$ . Since we obtained that  $k \ll \lambda_u$ , it follows that  $k' \sim 10\lambda_u$ , i.e. due to the overexpression of CasE, the processing rate becomes for an order of magnitude larger than pre-crRNA decay rate. Therefore, the overexpression of CasE makes the system switch from the state in which almost all of the generated pre-crRNA is degraded, to the state where most of the generated pre-crRNA is processed to crRNA.

We will below use the values of the system parameters that were estimated above, in order to numerically investigate kinetics of the transcript accumulation. To investigate the kinetics, we will simulate the system both deterministically and stochastically; we perform the stochastic simulations since the number of uninduced pre-crRNA and crRNA molecules are small, and since the number of pre-crRNA molecules becomes even smaller as CasE is overexpressed. However, we will see in the subsequent figures that the stochastic and deterministic results are in agreement with each other, which validates that the simple analytic expressions that we derive (e.g. Eq. (0.3)) can be used to describe the system.

We first numerically investigate how the amount of unprocessed and processed transcripts change as  $k$  is increased (i.e. as CasE is overexpressed). Stochastic simulations are performed by using Gillespie stochastic simulation algorithm [18], and stochastic trajectories are shown together with the deterministic curves. Figure 2A corresponds to the uninduced system, where the uninduced system parameters (see the previous section) lead to the experimentally observed steady state values ( $[u] \sim [p] \sim 10$ ). In Figure 2B, we increase the value of  $k$  1000 times; note that this  $k$  increase corresponds to





CasE overexpression in [11] (see above). We see that for such  $k$  increase  $[u]$  drops to a very small amount (few transcripts per cell), while  $[p]$  increases for about two orders of magnitude, consistently with the experimental observations. In Figure 2C,  $k$  is increased for an additional order of magnitude (i.e. 10000 fold relative to the uninduced value). This additional increase of  $k$  leads to even smaller amount of pre-crRNA, while the amount of crRNA increases for an additional small value (see the discussion below).

The results in Figures 2B and 2C clearly support Eq. (0.3). That is, in both of the panels, the steady-state amount of pre-crRNA decreases to very small levels ( $\Delta u \sim -10$ ), which leads to about two orders of magnitude increase of steady-state crRNA amount ( $\Delta p \sim 1000$ ). Note that this is in accordance with Eq. (0.3), given that the constant of proportionality between  $\Delta u$  and  $\Delta p$

equals 100 ( $\lambda_u/\lambda_p = 100$ ). Furthermore, both the decrease of  $\Delta u$ , and the increase of  $\Delta p$ , are somewhat larger in Figure 2C compared to Figure 2B, which is again consistent with the direct proportionality in Eq. (0.3). Therefore, both analytical and numerical results show that small pre-crRNA decrease leads to a large crRNA increase upon CasE over-overexpression. Interestingly, this strong amplification crucially depends on loss of pre-crRNA through fast non-specific degradation, i.e. on large  $\lambda_u/\lambda_p$  ratio (see Eq. (0.3)).

Furthermore, we note that the increase of  $k$  for one order of magnitude between Figures 2B and 2C, leads to only small additional increase of crRNA (relative to the one in Figure 2B), which we further refer to as saturation of crRNA upon increase of pre-crRNA processing rate. To additionally investigate this saturation, in Figure 3A we systematically predict the effect of  $k$  increase on

unprocessed (pre-crRNA) and processed (crRNA) transcript amounts. We see that, as  $k$  is increased beyond 1000 fold, the amounts of both pre-crRNA and crRNA reach saturation; i.e. pre-crRNA and crRNA amounts do not significantly change with further increase of  $k$ . The saturation value of crRNA increase corresponds to  $\sim 100$  fold.

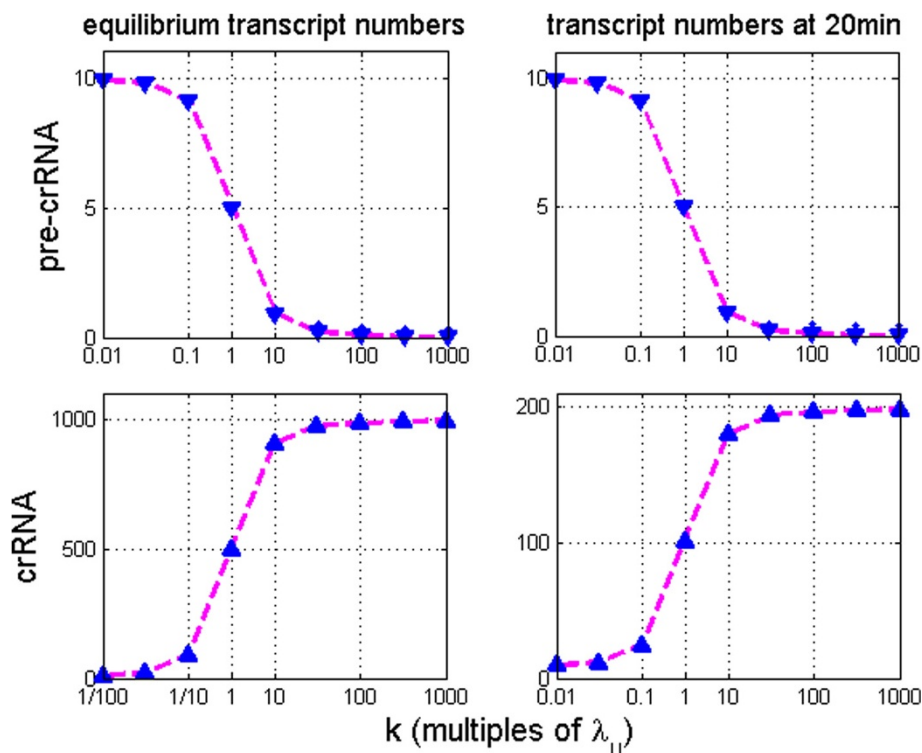
To analytically understand the observed saturation of crRNA upon  $k$  increase, it is straightforward to derive (see Methods) the relative increase of crRNA, as pre-crRNA processing rate is increased from  $k$  to  $k'$ :

$$\frac{\Delta[p]}{[p]} = \frac{\lambda_s/k + 1}{\lambda_s/k' + 1} - 1 \quad (0.5)$$

From the above equation, it follows that as  $k'$  becomes significantly larger than  $\lambda_s$  (i.e.  $k' \geq 10\lambda_s$ ),  $\Delta[p]/[p]$  no longer depends on  $k'$ .  $\Delta p/p$  then reaches saturation, i.e. approaches  $\lambda_s/k$ . Since  $\lambda_s \sim 100k$ , the saturation is reached when pre-crRNA processing rate is increased

for more than 1000 times, as a result of which  $\Delta[p]/[p]$  increases for about two orders of magnitude.

Finally, in Figure 3B, we investigate in more detail kinetics of crRNA accumulation. Figure 2 shows that the steady state is reached relatively slowly, i.e.  $\sim 300$  min after the system induction. However, when a virulent phage infects *E. coli*, the cell lysis is typically complete much before 300 min post-infection; e.g. for the well known *E. coli* T7 and T3 phages, the cell lysis starts at  $\sim 20$  min post-infection, with complete shot-off of host functions occurring much earlier [19]. Therefore, the steady state crRNA levels are likely not directly relevant for *E. coli* defense against phage infection. Due to this, in Figure 3B, we estimate crRNA levels at 20 min after the system induction. We see that, similarly to Figure 3A, as  $k$  is increased more than 1000 fold, crRNA amount at 20 min reaches saturation. While these saturation levels ( $\sim 200$  transcripts) are significantly smaller compared to the steady state values, they are still much larger than crRNA levels at which a partial protection against phage infection is observed ( $\sim 10$  crRNA transcripts as per [11]). Therefore, activation of



**Figure 3 Kinetics of crRNA accumulation.** The figure shows how pre-crRNA (the first row) and crRNA (the second row) amounts change as pre-crRNA processing rate ( $k$ ) is increased. CRISPR transcription rate remains constant and has the same value as in Figure 2. The first and the second column correspond, respectively, to **A)** equilibrium transcript amounts and **B)** transcript amounts at 20 min post-induction. The horizontal axes in the figure correspond to  $k$  in multiples of  $\lambda_u$ , where  $k$  changes from the uninduced value ( $\lambda_u/100$ ) to a very high value ( $1000\lambda_u$ ). The points on the horizontal axes are, for clearer presentation, plotted equidistantly, and correspond to  $k$  (in multiples of  $\lambda_u$ ) values of: (1/100, 1/50, 1/10, 1, 10, 50, 100, 500, 1000). The magenta line and the blue triangles correspond, respectively, to the stochastic and the deterministic simulations. The figure confirms the saturation effects observed in Figure 3, and suggests that the system is able to generate substantial crRNA amounts soon after its induction.

*cas* expression leads to a rapid accumulation of crRNA, which suggests such activation can lead to an effective protection against phage infection.

#### Joint overexpression of CRISPR and *cas* genes

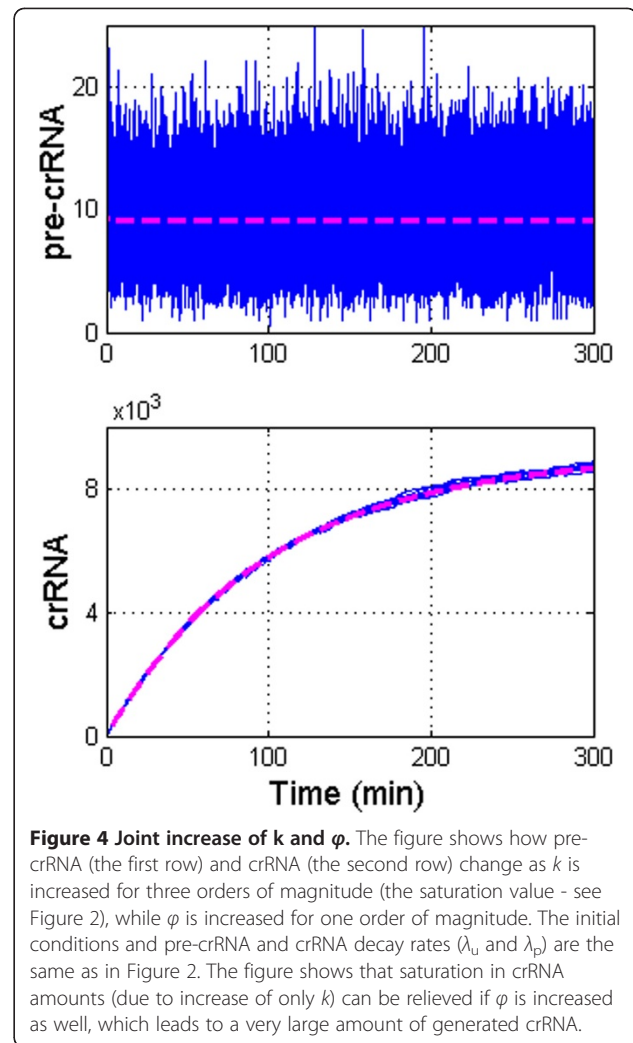
We next consider what happens if transcription of both *cas* genes and CRISPR array is activated. This analysis is, in part, motivated by reported repression of *cas* and (to a smaller extent) CRISPR promoters by H-NS, and by a model which proposes that the system is induced by abolishing this repression [10]. Activation of *cas* genes and CRISPR array transcription leads to increasing of both pre-crRNA processing rate (we assume that  $k$  increases to  $k'$ ) and CRISPR transcription rate (we assume that  $\phi$  increases to  $\phi'$ ). It is straightforward to derive (see Methods) that upon increase of both  $\phi$  and  $k$ , the amount of generated crRNA is given by:

$$\frac{\Delta[p]}{[p]} = \frac{(\lambda_s/k + 1) \phi'}{(\lambda_s/k' + 1) \phi} - 1 \quad (0.6)$$

From Eq. (0.6), we see that relative increase in crRNA depends linearly on relative increase of CRISPR transcription rate ( $\phi'/\phi$ ). From this follows that the saturation in crRNA due to increase of only  $k$ , which was discussed in the previous subsection, can be relieved if  $\phi$  is increased as well.

Increase of crRNA due to joint increase of  $k$  and  $\phi$  is numerically investigated in Figure 4. In this figure,  $k$  is increased for the amount that corresponds to the saturation (see the previous subsection), while  $\phi$  is increased tenfold. Note that the tenfold increase in  $\phi$  approaches maximal biochemically realistic value, since the basal  $\phi$  value is already moderately high ( $\sim 10 \text{ min}^{-1}$ ), while the transcription rate of very strong rRNA promoters is for about one order of magnitude higher [17]. We see that such induction strategy leads to an even higher increase in the amount of generated steady-state crRNA ( $\sim 10^3$  fold relative increase of crRNA upon induction); similarly, the amount of generated crRNA soon after the induction (e.g. at 20 min post-induction) - which may be relevant for defense against bacteriophages - is much higher than the minimal crRNA amount ( $\sim 10$  transcripts) necessary for partial protection against viruses [11].

In Figure 4, CRISPR transcription was increased in order to relieve the saturation due to increase of only  $k$  (compare with Figure 2C), and  $\phi$  was increased for a maximal biochemically realistic value. Consequently, crRNA amount in Figure 4 roughly corresponds to the maximal value that can be generated by the system. On the other hand, an increase in CRISPR transcription can be also used to substantially reduce the increase in pre-crRNA processing rate, while still achieving the same increase in generated crRNAs. This possibility is explored in Figure 5, which is,

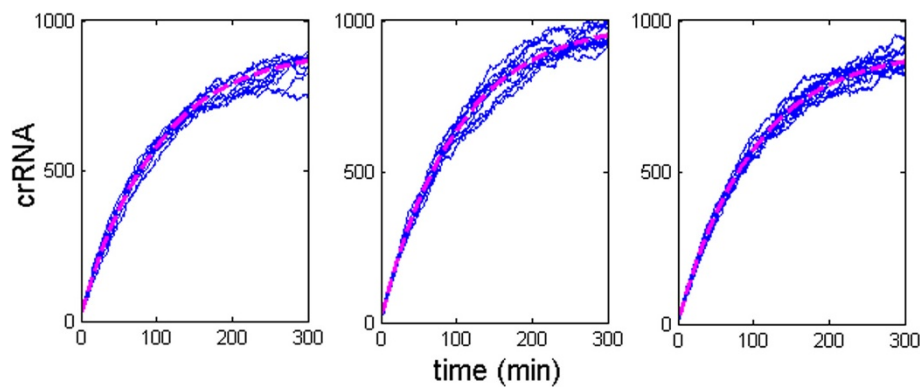


**Figure 4 Joint increase of  $k$  and  $\phi$ .** The figure shows how pre-crRNA (the first row) and crRNA (the second row) change as  $k$  is increased for three orders of magnitude (the saturation value - see Figure 2), while  $\phi$  is increased for one order of magnitude. The initial conditions and pre-crRNA and crRNA decay rates ( $\lambda_u$  and  $\lambda_p$ ) are the same as in Figure 2. The figure shows that saturation in crRNA amounts (due to increase of only  $k$ ) can be relieved if  $\phi$  is increased as well, which leads to a very large amount of generated crRNA.

in part, motivated by experiments in which H-NS repression of *cas* and CRISPR promoters is abolished [10,16]. Upon this abolishment, the amount of crRNA is increased for about two orders of magnitude, i.e. for the similar value as in CasE overexpression experiments [11,16].

Figure 5 demonstrates that the two orders of magnitude increase of crRNA can be achieved through very different levels of increase of pre-crRNA processing rate  $k$ , if CRISPR transcription rate is allowed to increase as well. Accordingly, the three panels in Figure 5, show roughly the same (two orders of magnitude) increase in crRNA levels, which are achieved in the following way; *i*) in Figure 5A,  $k$  is increased for three orders of magnitude, without increase of  $\phi$ , *ii*) in Figure 5B,  $k$  is increased for two orders of magnitude, while  $\phi$  is increased two times, *iii*) in Figure 5C, both  $k$  and  $\phi$  are increased for one order of magnitude.

Figure 5 demonstrates that large amounts of crRNA can be generated without a large CasE overexpression - which is characteristic for the (artificial) overexpression experiments - as long as CRISPR array transcription is increased



**Figure 5 Reducing the increase of  $k$  through increase of  $\phi$ .** The figure shows increase of crRNA as **A**) pre-crRNA processing rate ( $k$ ) is increased 1000 fold relative to the uninduced value, while CRISPR transcription rate ( $\phi$ ) is kept unchanged, **B**)  $k$  is increased 100 fold, while  $\phi$  is increased twofold, **C**) both  $k$  and  $\phi$  are increased 10 fold. The figure shows that a moderate increase in  $\phi$  allows to substantially reduce the increase in  $k$ , while still achieving the same increase in crRNA amount.

for a much smaller amount. While conditions of natural CRISPR/Cas induction are currently unclear [13], it is likely that activation of the CRISPR array promoter is much weaker compared to the activation of the *cas* promoter (note that repression of the *cas* promoter by H-NS was found to be significantly stronger than repression of the CRISPR promoter) [10,16]. This, therefore, suggests that conditions of natural system induction might roughly correspond to Figure 5B (the increase of  $k$  that is much larger than the increase of  $\phi$ ).

## Discussion

We here proposed a simple model of CRISPR transcript processing. We used this model, together with previous experimental measurements, to infer all the parameters that characterize the uninduced system. We showed that our model can explain the experimental observation that CasE-dependent decrease of very low initial steady-state level of *E. coli* pre-crRNA leads to a very large increase of crRNA abundance. Interestingly, this observation is a direct consequence of fast non-specific (i.e., not leading to crRNA) degradation of pre-crRNA. Our results, therefore, strongly suggest that non-specific degradation by an yet unidentified nuclease is a major control element of CRISPR expression and CRISPR/Cas response.

It is interesting to note that while effects of activation of *cas* gene transcription on CRISPR/Cas system were extensively studied, there is a lack of such studies for activation of CRISPR array transcription. Specifically, changes of pre-crRNA and crRNA amounts were quantitated only for *cas* gene overexpression, but not for CRISPR array overexpression [11]. Furthermore, while effects of *cas* gene overexpression on host protection against phage infection were measured [11], there is no such analysis for CRISPR array overexpression. That is, while in [12] it was shown that joint overexpression of *cas* genes and CRISPR array

leads to efficient protection against bacteriophage infection, it is unclear what additional protection is provided due to CRISPR array overexpression. Finally, activation by LeuO (a transcription regulator that abolishes H-NS repression) was studied for the *cas* promoter [16], but remains to be investigated for the CRISPR promoter.

Contrary to the almost complete emphasis on activation of *cas* gene transcription, the results presented here indicate that activation of CRISPR array transcription may be an important mechanism of CRISPR/Cas response. That is, we showed that there is a saturation of generated crRNA upon overexpression of only *cas* genes, i.e. that the amount of crRNA stops to increase when the rate of pre-crRNA processing is increased above certain level. This saturation is relieved when the rate of CRISPR transcription is increased as well, and we showed that a joint increase in transcription rates of *cas* and CRISPR promoters can lead to a very large (three orders of magnitude) increase of steady state crRNA levels. We, moreover, obtained that a substantial amount of crRNAs can be generated soon after the system induction, which suggests that the system may be capable for efficient protection against viruses under natural conditions. Unlike the situation observed in other bacteria, *E. coli* CRISPR spacers for the most part do no match sequences in known phages or plasmids. Yet, numerous data show that *E. coli* CRISPR/Cas system is functional once appropriate spacers are introduced by means of genetic engineering [12,20]. Presumably, the mechanism of CRISPR transcript processing, which was analyzed here, is relevant for protection against *E. coli* phages that are yet to be identified [21].

As a further support of potential importance of CRISPR array regulation, we showed that a modest increase of CRISPR transcription rate can substantially decrease for how much pre-crRNA processing rate needs to increase in order to achieve a desired crRNA amount.



For example, as small as twofold increase of CRISPR transcription rate allows reducing for one order of magnitude the pre-crRNA processing rate needed to achieve the two orders of magnitude increase of crRNAs (the increase observed when H-NS repression is abolished). Since repression of *cas* promoters by H-NS was found to be significantly stronger than the repression of CRISPR promoters, the regime in which the increase of pre-crRNA processing rate is significantly larger compared to the increase of CRISPR transcription rate may be directly relevant for natural system induction.

## Conclusions

We here developed a simple model of CRISPR transcript processing, and showed that this model is able to explain the existing experimental observations. The model shows that the relationship between the relevant biochemical quantities can be viewed as strong linear amplification, where this effect is a consequence of fast non-specific degradation of pre-crRNA. This implicates that the unidentified nuclease, which is responsible for the non-specific degradation, is a major control element of CRISPR/Cas response. We furthermore pointed to the potential importance of regulation of CRISPR array transcription, which may be another important mechanism of CRISPR/Cas system induction. Elucidating how the system is induced under natural conditions remains a major question to be addressed by both experimental and theoretical research.

## Methods

### Overexpression of *cas* genes

Kinetic equations that describe generation, degradation and processing of CRISPR transcripts (see Figure 1) are given below:

$$\frac{d[u]}{dt} = \phi - \lambda_u[u] - k[u] \quad (0.7)$$

$$\frac{d[p]}{dt} = -\lambda_p[p] + k[u] \quad (0.8)$$

Notation used in the above equations is described in Results and Figure 1. In the steady state  $d[u]/dt = 0$  and  $d[p]/dt = 0$ , so:

$$0 = \phi - \lambda_u[u] - k[u] \quad (0.9)$$

$$0 = -\lambda_p[p] + k[u] \quad (0.10)$$

Upon CasE overexpression, the new steady state becomes:

$$0 = \phi - \lambda_u[u]' - k'[u]' \quad (0.11)$$

$$0 = -\lambda_p[p]' + k'[u]' \quad (0.12)$$

In the above equations, note that upon CasE overexpression, CRISPR transcription rate  $\phi$  and crRNA stability

$\lambda_p$  do not change [11], while pre-crRNA processing rate increases to  $k'$ .

We next subtract Eq. (0.10) from Eq. (0.9) and subtract Eq. (0.12) from Eq. (0.11). We then again subtract these two expressions to obtain:

$$\lambda_u([u] - [u]') - \lambda_p([p]' - [p]) = 0$$

In the above expression,  $[u]' - [u]$  is the change of pre-crRNA amount upon CasE overexpression, which we label as  $\Delta[u]$ . Similarly, we label the change of crRNA as  $\Delta[p] = [p]' - [p]$ . We therefore have:

$$\Delta[p] = -\lambda_u/\lambda_p \Delta[u] \quad (0.13)$$

Furthermore, to calculate  $[u]/[u]'$ , we express  $[u]$  from Eq. (0.9) and  $[u]'$  from Eq. (0.11) to obtain:

$$\frac{[u]}{[u]'} = \frac{\lambda_u + k'}{\lambda_u + k} \sim 10 \quad (0.14)$$

Finally, we can solve for  $[p]'$  from Eqs. (0.11) and (0.12), and for  $[p]$  from Eqs. (0.9) and (0.10) to obtain:

$$\frac{\Delta[p]}{[p]} = \frac{\lambda_s/k + 1}{\lambda_s/k' + 1} - 1 \quad (0.15)$$

### Joint increase of *cas* and CRISPR transcription

When transcription of both CRISPR and *cas* genes is increased, we assume that CRISPR transcription rate increases from  $\phi$  to  $\phi'$ , while pre-crRNA processing rate increases from  $k$  to  $k'$ . Then Eq. (0.11) and Eq. (0.12) become:

$$0 = \phi' - \lambda_u[u]' - k'[u]' \quad (0.16)$$

$$0 = -\lambda_p[p]' + k'[u]' \quad (0.17)$$

After expressing  $[p]'$  from Eqs. (0.16) and (0.17), and  $[p]$  from Eqs. (0.9) and (0.10) we obtain:

$$\frac{\Delta[p]}{[p]} = \frac{\lambda_s/k + 1}{\lambda_s/k' + 1} \frac{\phi'}{\phi} - 1 \quad (0.18)$$

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

MD and KS conceived the work. MD and MD performed the analysis. MD wrote the paper, with the help of MD and KS. All authors read and approved the final manuscript.

### Reviewers' comments

Reviewer 1- Dr. Mikhail Gelfand

This is a good paper reporting a combination of experimental results and theoretical modeling. I have only number of editorial comments.

Authors' response

We thank Dr. Gelfand for positive comments. We have addressed the comments below.

Abstract, Background. Claiming that CRISPR repeats are palindromic is an overstatement. Firstly, the repeats are only approximately palindromic. Secondly, not all repeats have palindromic structure.

Authors' response

We now reworded the abstract so that the term palindromic no longer appears. This term was originally used in reference to CRISPR acronym (Clustered Regularly Interspaced Short Palindromic Repeats).

Abstract, Conclusions: "the extent of s gene activation" – it seems, "s" is all that was left from "casE".

Authors' response

This is now corrected.

The half-life sometimes is measured in "min", and sometimes, in "1/min".

Authors' response: This was a typo which we now corrected (the half-lives are now consistently measured in "min", while the decay rates are measured in "1/min".

Results, Model definition. Points (i) and (iv) partially repeat each other.

Authors' response

We now modified these two points so that the redundancy is removed.

Results, Uninduced system parameters: it would be better to state at the very beginning that k depends on the concentration of CasE.

Authors' response

We now included such statement before the beginning of this section (the third sentence in the second to last paragraph of the section Results, Model definition).

Results, Overexpression of cas genes, para. 2: "10 transcripts lead to two orders of magnitude increase" – is not a clear sentence.

Authors' response

The word decrease was missing from the sentence, we now corrected this. Using primes (k') is not good notation when kinetics is considered: at the first glance might get confused with the derivative.

Authors' response

We are aware that using primes is not an ideal choice; however, all alternative solutions (e.g. using subscripts) that we could think of are more cumbersome, and may potentially lead to confusion as well; note that not only k, but also values of all other quantities after the system induction are marked with primes ([U]', [P]',  $\phi'$ ). To prevent possible confusion, we now introduced the sentence in the text: "Note that primes in our notation correspond to the quantity values after the system induction, rather than to derivatives". We furthermore note that all the quantities in the text are clearly defined as they appear, which we hope prevents confusion with the derivatives.

Discussion, para. 1: the possibility to use CRISPR in synthetic circuits seems to be an overclaim.

Authors' response

We removed this statement, since this topic was indeed not analyzed in the paper.

Reviewer 2 - Dr. Eugen Koonin

CRISPR-Cas is an extremely intriguing system the regulation of which under different conditions remains poorly understood. Therefore, diverse approaches to this problem are of interest. Here the authors develop an extremely simple mathematical model to describe the effect of overexpression of cas genes, and in particular casE, on the production of crRNA. The results indicate that, when cas genes are over-expressed, the amount of mature crRNA increases proportionate to the decrease in the amount of pre-crRNA with a coefficient equal to the ratio of the degradation rates of pre-crRNA and crRNA. Because this ratio is large, about 100, there is the amplification effect that is the main claim of the article: a small decrease in the amount of pre-crRNA results in a large increase in the amount of crRNA. In other words, as the authors point out, the system switches from a 'non-productive' regime, when almost all pre-crRNA is degraded, to a 'productive' regime when almost all pre-crRNA is processed into crRNA. In more specific terms, this happens because pre-crRNA is unstable whereas crRNA is stable, so excess of Cas proteins prevents degradation of pre-crRNA, hence (nearly) all pre-crRNA molecules are channeled into crRNA resulting in the 'amplification effect'. I believe this scenario is straightforward and valid, and the model presented in the manuscript, although obvious, does quantify the effect, which makes the article worth the attention of researchers in the CRISPR field.

Authors' response

We thank Dr. Koonin for positive comments, and we addressed the suggestions regarding the paper presentation.

However, having asserted the above, I also believe that the manuscript is in need of serious rewriting. The current version is obscure to the point of being misleading. Although it is technically correct to call the CRISPR-Cas system a linear amplifier, in reality, I think the very term amplification has a major confusing potential. My first thought when reading the title of the paper was that this is about actual replication of crRNA. My suggestion would be to get rid of 'amplification' altogether and replace it with 'enhancement of crRNA production' or some such phrase.

Authors' response

We appreciate the comment that our statements - which were referring to relationships between biochemical quantities - may be confused for literal biological mechanisms, especially if a reader is exposed only to the title/abstract. To address this comment, we did the following: i) Removed the term "amplification" from the title, as suggested. ii) Rewrote the abstract so that it is now clear that "amplification" refers to the derived relationship between the relevant biochemical quantities, rather than to a literal biological mechanism. In particular, it is now explicitly stated in the abstract: "...The relationship between the decrease of pre-crRNAs and the increase of crRNAs corresponds to strong linear amplification...". Once the relevant explanation is provided, we did not completely remove the term amplification from the abstract since, as mentioned by Dr. Koonin, it accurately represents the obtained relationship between the relevant quantities. iii) In the main text the term "amplification" first appears in Introduction, but this is immediately after the sentence which makes clear that we are referring to the relationship between the relevant quantities. Given this explanation we kept the term amplification further in the text. "The model shows that the transcript processing corresponds to strong linear amplification of pre-crRNA. The strong amplification is due to fast non-specific degradation by an unspecified nuclease, suggesting that this nuclease is a major control element of CRISPR response".

The first sentence in this quote is simply not understandable. The second one creates the impression that the uncharacterized nuclease somehow directly promotes that 'amplification'. The reality is that degradation of pre-crRNA as such leads to low level of crRNA production, just as one would expect; it is the prevention of the said degradation that produces the observed effect. There is much more such language in the current manuscript, so I think it should be carefully re-read and edited/rewritten.

Authors' response

Actually, the meaning of these statements was that the 'amplification' directly depends on to the ratio of the decay rates of pre-crRNA and crRNA (as mentioned above by Dr. Koonin). Since the large decay rate of pre-crRNA is due to fast non-specific processing by an unidentified nuclease, this nuclease is likely a major control element of CRISPR response. However, similarly as above, we appreciate that the statements can be confused with the endonuclease being directly physically involved in crRNA amplification. We consequently removed/rewrote the two sentences, and the term "amplification" is now completely absent from Abstract-Conclusion. We also rewrote the relevant sentences in the main text along the same lines, so as to avoid possible confusion.

Reviewer 3 - Dr. L. Aravind

The large amplification of the crRNAs with the concomitant decline of pre-crRNAs to near undetectability upon CasE overexpression is an interesting factor in the action of the CRISPR/Cas system. The authors attempt to explain this with a mathematical model and also suggest that their model might be useful to understand endogenous CRISPR processing. One useful point the authors show is that although the promoter of the CRISPR locus in *E.coli* has been described as weak, it is mainly a consequence of degradation of pre-crRNA rather its low transcription. Then using their model they explain quite clearly how the observed amplification can be accounted for. Their model also shows large increase in number of crRNAs is attained at distinct increased rates of pre-crRNA processing when combined with CRISPR transcription rate increase. At least some of these appear to provide reasonable pairs of values which might resemble physiological induction.

Authors' response

We thank Dr. L. Aravind for positive comments, and we have addressed the suggestions below.

While the model is well-explained and appears to account for the observations published earlier by the authors, a question remains regarding its actual significance for the *E.coli* CRISPR/Cas system: The system does not appear to be induced at by PRD1 or lambda for that matter. Further the

endogenous CRISPR sequences of *E. coli* do not seem to match any apparently prevalent plasmids of viruses. The authors state: "We, moreover, obtained that a substantial amount of crRNAs can be generated soon after the system induction, which suggests that the system may be capable for efficient protection against viruses under natural conditions." While in principle this is correct, the actual situation seems to be different. If the Cas and CRISPR genes are not induced upon phage infection in *E. coli* then the high crRNA levels shortly after induction becomes moot. So it might be useful for the authors to stress that this aspect of their model is relevant only if the CRISPR/Cas genes are induced as they propose by some hypothetical phage (i.e. as yet unknown *E. coli* invasive DNA), even though this has not been actually observed in *E. coli* with any of the currently studied phages. Authors' response

To address this comment we added the last three sentences in the second to the last paragraph of the Discussion. As discussed by Dr. Aravind, *E. coli* CRISPR spacers indeed do not match sequences of known *E. coli* phages or plasmids. However, numerous data show that once appropriate spacers are introduced in CRISPR array by means of genetic engineering, *E. coli* CRISPR/Cas system becomes functional. Consequently, it is currently mostly assumed that CRISPR/Cas system should indeed be induced by invasive DNA (see e.g. a model summarized by Figure 7 in [10]), though exact physiological conditions under which such induction occurs have yet to be understood. Furthermore, it is estimated that a very large proportion (in fact, most) of *E. coli* phages are not yet known [21], which may provide a reason for the absence of a match between *E. coli* CRISPR spacers and sequences of known phages. Finally, our prediction that the system can generate high crRNA levels shortly after its induction might prove to be relevant even if some (currently) unexpected function of CRISPR/Cas system in *E. coli* emerges.

Minor issues:

Check spellings/unclear expressions:

"archaeal"

"the processing rate becomes for an order of magnitude larger than the transcription decay rate"

It appears that the authors are talking about unprocessed RNA decay rate.

Brackets might help presentation of equation 1.6.

"observed large generation of crRNAs from only few pre-crRNAs" observed amplification. . .

Authors' response

We corrected/implemented all the suggested changes in the text.

## Acknowledgements

This work is supported by a Marie Curie International Reintegration Grant within the 7<sup>th</sup> European Community Framework Programme (PIRG08-GA-2010-276996) and by the Ministry of Education and Science of the Republic of Serbia, under project No. ON173052. Magdalena Djordjevic is supported in part by a Marie Curie International Reintegration Grant within the 7<sup>th</sup> European Community Framework Programme (PIRG08-GA-2010-276913). KS acknowledges support through NIH Grant GM59295, Russian Academy Presidium Program in Molecular And Cell Biology and a Russian Foundation for Basic Research grant.

## Author details

<sup>1</sup>Institute of Physiology and Biochemistry, Faculty of Biology, University of Belgrade, Belgrade, Serbia. <sup>2</sup>Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia. <sup>3</sup>Institutes of Molecular Genetics and Gene Biology, Russian Academy of Sciences, Moscow, Russia. <sup>4</sup>Department of Molecular Biology and Biochemistry, Rutgers, State University of New Jersey, Piscataway, USA. <sup>5</sup>Waksman Institute, Rutgers, State University of New Jersey, Piscataway, USA.

Received: 10 April 2012 Accepted: 6 July 2012

Published: 31 July 2012

## References

- Mojica FJ, Diez-Villasenor C, Soria E, Juez G: **Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.** *Mol Microbiol* 2000, **36**:244–246.

- Jansen R, Embden JD, Gaastra W, Schouls LM: **Identification of genes that are associated with DNA repeats in prokaryotes.** *Mol Microbiol* 2002, **43**:1565–1575.
- Horvath P, Barrangou R: **CRISPR/Cas, the immune system of bacteria and archaea.** *Science* 2010, **327**:167–170 (New York, NY).
- Jansen R, van Embden JD, Gaastra W, Schouls LM: **Identification of a novel family of sequence repeats among prokaryotes.** *Omnics: a journal of integrative biology* 2002, **6**:23–33.
- Bolotin A, Quinquis B, Sorokin A, Ehrlich SD: **Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.** *Microbiology* 2005, **151**:2551–2561 (Reading, England).
- Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E: **Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.** *J Mol Evol* 2005, **60**:174–182.
- Pourcel C, Salvignol G, Vergnaud G: **CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies.** *Microbiology* 2005, **151**:653–663. Reading, England.
- Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV: **A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action.** *Biology direct* 2006, **1**:7.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P: **CRISPR provides acquired resistance against viruses in prokaryotes.** *Science* 2007, **315**:1709–1712. New York, NY.
- Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R: **Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS.** *Mol Microbiol* 2010, **75**:1495–1512.
- Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K: **Transcription, processing and function of CRISPR cassettes in *Escherichia coli*.** *Mol Microbiol* 2010, **77**:1367–1379.
- Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J: **Small CRISPR RNAs guide antiviral defense in prokaryotes.** *Science* 2008, **321**:960–964. New York, NY.
- Al-Attar S, Westra ER, van der Oost J, Brouns SJ: **Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes.** *Biol Chem* 2011, **392**:277–289.
- Diez-Villasenor C, Almendros C, Garcia-Martinez J, Mojica FJ: **Diversity of CRISPR loci in *Escherichia coli*.** *Microbiology* 2010, **156**:1351–1361. Reading, England.
- Poranen MM, Ravaniti JJ, Grahn AM, Gupta R, Auvinen P, Bamford DH: **Global changes in cellular gene expression during bacteriophage PRD1 infection.** *J Virol* 2006, **80**:8081–8088.
- Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, Stratmann T, Wurm R, Raine A, Mescher M, Van Heereveld L, et al: **H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator *LeuO*.** *Mol Microbiol* 2010, **77**:1380–1393.
- Sneppen K: *GZ: Physics in Molecular Biology*. Cambridge: Cambridge University Press; 2005.
- Gillespie DT: **Stochastic simulation of chemical kinetics.** *Annu Rev Phys Chem* 2007, **58**:35–55.
- Kruger DH, Schroeder C: **Bacteriophage T3 and bacteriophage T7 virus-host cell interactions.** *Microbiol Rev* 1981, **45**:9–51.
- Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K: **Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence.** *Proc Natl Acad Sci U S A* 2011, **108**:10098–10103.
- Semenova E, Nagornykh M, Pyatnitskiy M, Artamonova II, Severinov K: **Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*.** *FEMS Microbiol Lett* 2009, **296**:110–116.

doi:10.1186/1745-6150-7-24

Cite this article as: Djordjevic et al.: CRISPR transcript processing: a mechanism for generating a large number of small interfering RNAs. *Biology Direct* 2012 **7**:24.